

周报（2014.04.21-2014.04.27）

1. 寻找并查看网上已有的开源的 **data integration** 的代码
2. 从网络上爬取数据。已经完成输入网址获得网页的内容。当一个网页是多个新闻的列表时，需要分析网页的内容，以获得所有完整的新闻数据，这部分目前没有实现，打算下周一周二实现。

下周工作：

1. 完成网络爬虫的工作。
2. 了解开源的 **data integration** 的代码，了解大概的流程。